ASSESSING IMAGE FUSION METHODS FOR UNCONSTRAINED OUTDOOR SCENES

J. Zeng*, A. Sayedelahl, T. Gilmore, P. Frazier and M. Chouikha
Department of Electrical and Computer Engineering
Howard University, Washington, DC 20059

ABSTRACT

In this paper, we assess nine recent pixel-level image fusion algorithms. These fusion algorithms are experimentally evaluated with quantitative assessment techniques. Finally, a new assessment paradigm for image fusion is provided.

1. INTRODUCTION

Image fusion integrates images of the same target or scene from multiple sensors to produce a composite image or images that will inherit most salient features from the individual images. A resulting fused image can be thought of as an image taken by an advanced not-yet-existing imaging sensor that can capture multiple salient features of the scene simultaneously. The fused image usually has more information about the target or scene than any of the individual images used in the fusion process. Images used for fusion can be taken from multi-modal imaging sensors or from the same imaging sensor at different times. The target or scene in the images can be exactly the same or partially the same (e.g., images taken from slightly different angles, some objects in some images partially occluded or disappeared, or new objects added to the scene).

Image fusion has been investigated by many research groups and a number of algorithms have been developed (Zhang and Blum, 1999; Scheunders and De Backer, 2001; Rajan and Chaudhuri, 2002; Chan et al. 2003). Although each algorithm has shown some promising aspects, there seems to be a lack of universal criteria to measure the quality of the fusion algorithms. In many cases, qualitative criteria such as visual analysis are used to assess the resulting fused images (Toet and Franken, 2003). Recently, some quantitative measures have been developed (Xydeas and Petrovic, 2000; Qu et al., 2002; Piella and Heijmans, 2003; Wang et al. 2003; Wang et al., 2004).

Efforts have been made to review image fusion techniques and assess their qualities (Valet et al. 2001; Piella, 2003; Smith and Heather, 2005; Sadjadi, 2005), but there are still many open issues to be resolved in this area, especially in terms of image fusion assessment. This paper aims at providing some new perspectives in assessment of fusion algorithms and filling the gap between theoretical and practical assessments. In addition, it provides a new fusion assessment paradigm in order to bridge the gap between theoretically sound but impractical assessments and practically sound but theoretically unproven assessments. We focus on unconstrained outdoor scenes because of the targeted applications of this research. Unconstrained outdoor scenes generally refer to natural suburban or rural outdoor scenes (e.g., terrains and mountainous areas). They can also include urban outdoor scenes (e.g., roads and buildings) as they naturally appear. These outdoor scenes tend to have more background noises caused by the environmental and weather factors. Distances to the targets also tend to be greater than indoor scenes, and as a result, the targets usually appear smaller and less clear in the images. These will thus make image fusion more difficult, but applications in such scenes can also benefit more from the multisensor image fusion.

2. IMAGE FUSION ALGORITHMS

2.1 Image Fusion Definition

Currently, there is no universally accepted definition of image fusion, but the objective of image fusion is clear - improved quality of information gained from the fused image. Image fusion can thus be simply put as a framework where a composite image (or images) can be produced that contains improved quality of information about the target or scene compared to individual source images. All sources

Report Documentation Page

Form Approved
OMB No. 0704-0188

1. REPORT DATE **01 NOV 2006**	2. REPORT TYPE **N/A**	3. DATES COVERED **-**

4. TITLE AND SUBTITLE **Assessing Image Fusion Methods For Unconstrained Outdoor Scenes**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Department of Electrical and Computer Engineering Howard University, Washington, DC 20059**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S) 11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release, distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM002075.

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**UU**	**7**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

images have to be aligned or registered before being fused.

2.2 Overview of Image Fusion Algorithms

2.2.1 Fusion using principle component analysis (PCA)

The PCA-based fusion method simply uses pixel values of all source images at each pixel location, adds a weight factor to each pixel value, and takes an average of the weighted pixel values as the result for the fused image at the same pixel location. The optimal weight factors are determined by the PCA technique.

2.2.2 Fusion using Laplacian pyramid

The Laplacian pyramid fusion consists of an iterative process of calculating Gaussian and Laplacian pyramids of each source image, fusing the Laplacian images at each pyramid level by selecting the pixel with larger absolute values, combining the fused Laplacian pyramid with the combined pyramid expanded from the lower level, and expanding the combined pyramids to the upper level. The pixel selection step above can also be done using a PCA-based weighted averaging technique.

2.2.3 Fusion using ratio of low pass pyramid (RoLP)

In the above Laplacian pyramid method, simply replace the difference (i.e., the Laplacian pyramid) with a division operation to obtain the RoLP pyramids.

2.2.4 Fusion using contrast pyramid

In the above RoLP pyramid method, simply replace the division with the following contrast formula to obtain the contrast pyramids:

$$Con_k = \frac{G_k - EXPAND(G_{k+1})}{EXPAND(G_{k+1})}$$

where Con_k represents the contrast between two successive levels G_k and G_{k+1} in the Gaussian pyramids, and operation '*EXPAND*' consists of a simple upsampling followed by a low-pass filtering.

2.2.5 Fusion using gradient pyramid

A gradient pyramid is obtained by applying a set of 4 directional gradient filters (horizontal, vertical and 2 diagonal) to the Gaussian pyramid at each level. At each level, these 4 directional gradient pyramids are combined together to obtain a combined gradient pyramid that is similar to a Laplacian pyramid. The gradient pyramid fusion is therefore the same as the fusion using Laplacian pyramid except replacing the Laplacian pyramid with the combined gradient pyramid.

2.2.6 Fusion using filter-subtract-decimate (FSD) pyramid

The FSD pyramid fusion is conceptually identical to the Laplacian pyramid fusion method. The only difference is in the step of obtaining the difference images in creating the pyramid. In Laplacian pyramid, the difference image L_k at level k is obtained by subtracting an image upsampled and then low-pass filtered at level $k+1$ from the Gaussian image G_k at level k, while in FSD pyramid, this difference image is obtained directly from the Gaussian image Gk at level k subtracted by the low-pass filtered image of Gk. As a result, FSD pyramid fusion method is computationally more efficient than the Laplacian pyramid method by skipping an upsampling step.

2.2.7 Fusion using morphological pyramid

A morphological pyramid is obtained by applying morphological filters to the Gaussian pyramid at each level and taking the difference between 2 neighboring levels. A morphological filer is usually for noise removal and image smoothing. It is similar to the effect of a low-pass filter, but it does not alter shapes and locations of objects in the image. The morphological pyramid fusion is therefore the same as the fusion using Laplacian pyramid method except replacing the Laplacian pyramid with the morphological pyramid.

2.2.8 Fusion using discrete wavelet transform (DWT)

In the DWT-based fusion, the source images are first transformed by DWT to their corresponding wavelet coefficient images at each scale level. Corresponding approximation coefficients and detail coefficients of the source images at each level are then fused, respectively, based on a certain fusion rule.

This rule can be a simple addition or averaging, or it can be a PCA-based weighted averaging. The fused approximation and detail coefficients at each level are used in the final reconstruction of a single output fused image by an inverse DWT.

2.2.9 Fusion using Harr DWT method for shift invariance

One of the shortcomings of DWT method is its shift-variance which means that the DWT result varies if the source images are shifted, even slightly. Shift variance is caused by the downsampling step in the image decomposition process of the DWT. The Harr DWT is a shift-invariant DWT (SIDWT) that can solve the above problem by skipping the downsampling step in the decomposition process and using a new set of filters at each decomposition level.

3. PERFORMANCE ASSESSMENT OF IMAGE FUSION ALGORITHMS

Assessment of image fusion performance can be first divided into two categories: one with and one without reference images. In reference-based assessment, a fused image is evaluated against the reference image which serves as a ground truth. In assessment without reference images, the fused images are evaluated against the original source images for similarity. Furthermore, fusion assessment can be classified as either qualitative or quantitative in nature. In practical applications, however, neither qualitative nor quantitative assessment alone will satisfy the needs perfectly. Given the nature of complexity of specific applications, a new assessment paradigm combing both qualitative and quantitative assessment will be most appropriate in order to achieve the best assessment result.

3.1 Qualitative Assessment

Visual analysis and statistical analysis are the major forms of qualitative assessment for image fusion (Toet and Franken, 2003; Smith and Heather, 2005). In qualitative assessment, user performance is usually used as an indirect criterion to assess performance of image fusion in a specific task. For example, Receiver Operating Characteristic (ROC) analysis, which is widely used especially in medical imaging applications, has been used to evaluate medical image fusion (Twellmann et al., 2004; Zhou et al., 2002). Designed to describe the accuracy of diagnostic tests by a radiologist, ROC plots a curve between true positive fraction vs. false positive fraction in a series of diagnostic tests. It has now been expanded to many other applications where ground truth is readily available. Other examples of qualitative assessment include Measure of Performance (MoP), which measures how much a fusion technique helps in getting the job done (e.g., workload reduction) (Smith and Heather, 2005).

3.2 Quantitative Assessment

3.2.1 Evaluation with reference images

A commonly used reference-based assessment metric is the root mean square error (RMSE) which is defined as follows:

$$RMSE = \sqrt{\tfrac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}(R(m,n)-F(m,n))^2}$$

where *R(m,n)* and *F(m,n)* are reference and fused images, respectively, and M and N are image dimensions.

3.2.2 Evaluation without reference images

In many practical applications, reference images are not readily available. A number of quantitative metrics have been investigated to assess image fusion in these situations.

<u>**Mutual information (Qu et al., 2002)**</u>

Mutual information (MI) is defined as follows.

$$I(A,B) = \sum_{a,b} P_{AB}(a,b)\log\tfrac{P_{AB}(a,b)}{P_A(a)P_B(b)}$$

where $P_{AB}(a,b)$ is the joint distribution probability, $P_A(a)$ and $P_B(b)$ are the distribution probabilities of A and B, respectively. The mutual information I(A,B) measures similarity of image intensity distribution between images A and B. Distribution probabilities can be obtained using image histograms.

When reference images are available, the following steps can be used to assess the fusion algorithms with MI:

a. Calculate mutual information between fused image and the reference image I(*F*,*R*).
b. A higher value of I(*F*,*R*) indicates better similarity between *F* and *R*, and thus a better fusion algorithm.

When no reference images are available, fusion assessment is performed as follows:

a. The MI-based measure is defined as

$$M_F(A,B)=I(F,A)+I(F,B)$$

b. $M_F(A,B)$ represents total amount of similarity between fused image F vs. source images A and B. Again, a higher value indicates a better fusion algorithm.

Structural Similarity Index (Wang et al., 2004)

Focused on the link between the structural information changes in images and the perceived distortions of the images, the Structural Similarity (SSIM) index is defined as a measure to assess similarity of two images A and B as follows:

$$SSIM(A,B)=\frac{(2\mu_A\mu_B+C_1)(2\sigma_{AB}+C_2)}{(\mu_A^2+\mu_B^2+C_1)(\sigma_A^2+\sigma_B^2+C_2)}$$

where μ_A and μ_B are the mean intensities of images A and B, and σ_A and σ_B are standard deviations of A and B, respectively. σ_{AB} is the covariance of A and B, and C1 and C2 are small constants for A and B, respectively. A higher value of SSIM index corresponds to greater similarity of the two images. As with the mutual information, performance of image fusion can be assessed by the following total amount of similarity between fused image F vs. source images A and B:

$$M_F(A,B)=SSIM(F,A)+SSIM(F,B)$$

A higher value of $M_F(A,B)$ corresponds to a better fusion algorithm. A revised similarity measurement was also proposed based on a simplified version of SSIM (Piella and Heijmans, 2003).

Perceptual edge preservation value (Xydeas and Petrovic, 2000)

The perceptual edge preservation metric emphasizes the importance of edge information in image fusion quality as follows:

a. It associates important visual information with 'edge' information, similar to human visual system.
b. It measures the amount of edge information that is transferred from the source images to the fused image.

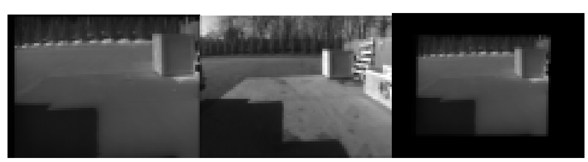
Other assessment metrics

A number of other non-reference quantitative fusion assessment metrics have also been developed. They include Measures of Effectiveness (MOE) aiming at the separation of target region and background region in the fused image (Sadjadi, 2005), an assessment metric based on the amount of retention of individual sensor information (e.g., visible, thermal) in the fused images (Ulug and McCullough, 2000), and a quantitative assessment of the signal-level image fusion with the presence of noses in source images (Petrovic and Xydeas, 2003).

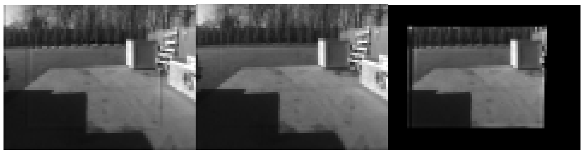
4. EXPERIMENTAL RESULTS

Two pairs of images (one infrared and one visible camera images in each pair) are used to experimentally evaluate the nine fusion algorithms described in 2.2. The images are first registered using control point mapping registration technique. Figgures 1 and 2 show the fusion results of the two pairs of images.

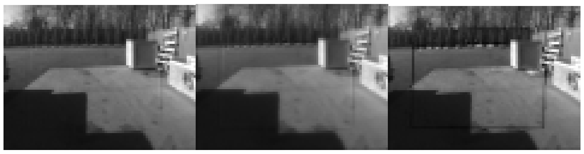
In order to assess the fusion algorithms, we have applied the mutual information (MI) method and structural similarity index (SSIM) method to all the fused images. The assessment results, ordered in descending quality rank each, are shown in Tables 1 through 4.



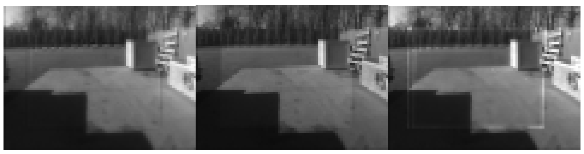
Original-IR Original-visible Registered image



Fused-Laplacian Fused-DWT Fused-Cont. Pyr.



Fused-Harr DWT Fused-Grd. Pyr. Fused-Morp. Pyr.



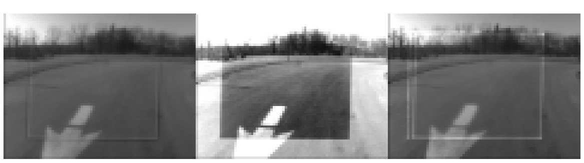
Fused-FSD Pyr. Fused-PCA Fused-RoLP Pyr.

Fig. 1 Experimental results of fusion algorithms with the image 'Parking Lot'

Original–IR Original–visible Registered image

Fused-Laplacian Fused-DWT Fused-Cont. Pyr.

Fused-Harr DWT Fused-Grd. Pyr. Fused-Morp. Pyr.



Fused–FSD Pyr. Fused–PCA Fused–RoLP Pyr.

Fig. 2 Experimental results of fusion algorithms with the image 'Road'

Tab. 1 MI for the image 'Parking Lot'

Rank	Algorithms	$MI_{(F/V)}$	$MI_{(F/IR)}$	Sum
1	FSD	88.151	73.094	161.245
2	Gradient	88.15	73.089	161.239
3	RoLP	83.166	73.316	156.482
4	Morphology	82.169	73.266	155.435
5	DWT	80.31	73.421	153.731
6	Laplacian	79.831	73.447	153.278
7	HarrDWT	79.516	73.272	152.788
8	PCA	81.397	70.293	151.69
9	Contrast	70.435	43.23	113.665

Tab. 2 MI for the image 'Road'

Rank	Algorithms	$MI_{(F/V)}$	$MI_{(F/IR)}$	Sum
1	Harr DWT	86.105	62.193	148.298
2	FSD	80.666	66.967	147.633
3	Gradient	80.622	66.974	147.596
4	Laplacian	78.475	67.551	146.026
5	RoLP	78.38	67.536	145.916
6	Morphology	75.261	67.673	142.934
7	DWT	74.098	67.456	141.554
8	PCA	78.475	58.514	136.989
9	Contrast	66.304	39.745	106.049

Tab. 3 SSIM index for the image 'Parking Lot'

Rank	Algorithms	$SSIM_{(F/V)}$	$SSIM_{(F/IR)}$	Sum
1	PCA	0.96668	0.38303	1.34971
2	Gradient	0.95483	0.38266	1.33749
3	FSD	0.95373	0.38152	1.33525
4	Harr DWT	0.96657	0.36518	1.33175
5	Laplacian	0.9591	0.36869	1.32779
6	DWT	0.96127	0.35837	1.31964
7	RoLP	0.92238	0.38555	1.30793
8	Morphology	0.92577	0.36701	1.29278
9	Contrast	0.44046	0.83744	1.2779

Tab. 4 SSIM index for the image 'Road'

Rank	Algorithms	$SSIM_{(F/V)}$	$SSIM_{(F/IR)}$	Sum
1	Gradient	0.97014	0.39633	1.36647
2	FSD	0.96918	0.39567	1.36485
3	Laplacian	0.97174	0.38161	1.35335
4	DWT	0.97297	0.37855	1.35152
5	Morphology	0.9527	0.38615	1.33885
6	RoLP	0.90961	0.40215	1.31176
7	Contrast	0.46194	0.82889	1.29083
8	Harr DWT	0.85926	0.40698	1.26624
9	PCA	0.66323	0.18719	0.85042

5. DISCUSSIONS

It is noted from the assessment results that there is not a single fusion algorithm that can consistently rank on top of the other algorithms. This may suggest that performance of image fusion algorithms depends on images of specific applications. On the other hand, we also notice that some fusion algorithms do consistently rank high, such as gradient pyramids and FSD pyramids methods. It may hint that these algorithms are better tuned and more robust although more tests are needed to verify.

Although it is not totally convincing that fusion performance can be measured solely by the MI or SSIM values, especially when the values are close to each other, they do provide a reasonably good quantitative metric. In real applications, information from different sensors is not likely to be treated equally important. That is, information from some sensors is more emphasized than information from other sensors. For example, during nighttime battlefield missions, infrared sensors may be relied more heavily than optical cameras. Furthermore, since fused images are used to enhance visual information for human users, performance assessment of image fusion should be first judged by the users based on the mission of specific applications. Quantitative measures should only serve as a useful tool to assist human users to make difficult judgments whenever necessary.

As such, we introduce a weighted measure as a new quantitative assessment. Let W_A and W_B be the weights for images A and B with default values of $W_A=W_B=0.5$, and F be the fused image. Define the new weighted measure as follows:

$$M_F(A,B) = W_A J(F,A) + W_B J(F,B)$$

where J can be MI, SSIM or any other similarity metrics. Weights W_A and W_B are assigned different values based on their importance in specific applications. For example, their values can be set to 5 levels: 0.7 – 0.9 or more = most important; 0.5 – 0.7 = very important; 0.3 – 0.5 = important; 0.1 - 0.3 = less important; and 0.1 or less = not important. Also, they should always follow the constraint $W_A+W_B=1$. Here, larger value of $M_F(A,B)$ implies a better image fusion quality.

A new fusion assessment paradigm is thus proposed as follows:

a. First, let a human user use qualitative measures (e.g., visual analysis, statistical analysis, ROC, MoP) to make initial judgments about the fused images obtained with various fusion methods.
b. Second, use quantitative assessment measures, including the above weighted measure, to verify his/her judgments, or to provide assistance for a judgment if the difference between fused images is not qualitatively clear.
c. The above steps are repeated until a satisfactory judgment is reached. The final judgment is solely on the user side.

6. CONCLUSIONS

Image fusion techniques have shown some good progress in recent years. They are expected to play significant roles in many applications, especially in military such as US Army Future Combat Systems (FCS). This paper reviewed some of the latest image fusion algorithms and their performance assessment techniques. These fusion algorithms were applied to some outdoor images in the experiment and they showed mixed performance during assessment. We have come to the conclusion that no fusion algorithm always outperforms the others and performance of a fusion algorithm relies on images of specific applications. A combined qualitative and quantitative assessment approach seems to be the best way to determine which fusion algorithm is most appropriate for an application. Quantitative assessment metrics, however, cannot dominate the decision process. Instead, they are best served as a tool to provide the user with some additional technical evidence for his/her decision-making.

Despite recent technical progress, image fusion has yet reached the stage where it can be reliably and massively applied to a broad range of applications. More research is needed to develop more robust fusion algorithms and related hardware devices for real-time practical use. For example, combination of image fusion techniques at all levels (i.e., signal-level, pixel-level, feature-level, and symbolic level) will be likely to improve the quality of fused images and thus lead to improved decision-making. Other useful image features that have not been made full use of, such as color, will need to be further explored for image fusion purpose (Smith and Heather, 2005; Fay et al., 2000). At the same time, performance assessment of image fusion should continue to be shared between qualitative and quantitative methods, with increasing weight being placed toward new quantitative assessment techniques. Human users will continue to be the sole final decision-makers while improved image fusion techniques will continue to relieve their workloads and help them make quicker and more accurate decisions.

ACKNOWLEDGEMENTS

This research was supported, in part, by the US Army Robotics CTA Program.

REFERENCES

Chan, L. A., Der, S. Z., and Nasrabadi, N. M., 2003: Dualband FLIR fusion for automatic target recognition, Information Fusion, 4, 35-45.

Fay, D. A., Waxman, A. M., Aguilar, M., Ireland, D. B., Racamato, J. P., Ross, W. D., Streilein, W. W., Braun, M. I., 2000: Fusion of multisensor imagery for night vision: Color visualization, target learning and search, Third International Conference on Information Fusion, Paris.

Petrovic, V. S. and Xydeas, C. S., 2003: Sensor noise effects on signal-level image fusion performance, Information Fusion, 4, 167-183.

Piella, G., 2003: A general framework for multiresolution image fusion: from pixel to regions, Information Fusion, 4, 259–280.

Piella, G. and Heijmans, H., 2003: A new quality metric for image fusion, Proc. IEEE International Conference on Image Processing.

Qu, G., Zhang, D., and Yan, P., 2002: Information measure for performance of image fusion, Electronics Letters, Vol. 38, No. 7, pp. 313-315.

Rajan, D. and Chaudhuri, S., 2002: Data fusion techniques for super-resolution imaging, Information Fusion, 3, 25-38.

Sadjadi, F., 2005: Comparative Image Fusion Analysis, Proc. Second Joint IEEE International Workshop on Object Tracking and Classification in and Beyond the Visible Spectrum (OTCBVS'05).

Scheunders, P. and De Backer, S., 2001: Multispectral image fusion and merging using multiscale fundamental forms, Proc. IEEE International Conference on Image Processing.

Smith, M. and Heather, J., 2005: Review of image fusion technology in 2005, www.waterfallsolutions.co.uk,.

Toet, A. and Franken, E. M., 2003: Perceptual evaluation of different image fusion schemes, Displays, 24(1): 25–37.

Twellmann, T., Saalbach, A., Gerstung, O., Leach, M. O., and Nattkemper, T. W., 2004: Image fusion for dynamic contrast enhanced magnetic resonance imaging, BioMedical Engineering OnLine 2004, 3:35, www.biomedical-engineering-online.com/content/3/1/35

Ulug, M. E. and McCullough, C. L., 2000: A quantitative metric for comparison of night vision fusion algorithms, Sensor Fusion: Architectures, Algorithms, and Applications IV, SPIE Vol. 4051, 80-88.

Valet, L., Mauris, G., and Bolon, P., 2001: A statistical overview of recent literature in information fusion, IEEE AESS System Magazine..

Wang, Q., Shen, Y., Zhang, Y., and Zhang, J. Q., 2003: A quantitative method for evaluating the performances of hyperspectral image fusion, IEEE Transactions on Instrumentation and Measurement, Vol. 52, No. 4.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., 2004: Image quality assessment: From error visibility to structural similarity, IEEE Transactions on Image Processing, Vol. 13, No. 4.

Xydeas, C. S. and Petrovic, V., 2000: Objective image fusion performance measure, Electronics Letters, Volume: 36, Issue 4, pp. 308-309.

Zhang, Z. and Blum, R. S., 1999: A Categrorization of Multiscale-decomposition-based Image Fusion Schemes with a Performance Study for a Digital Camera Application", Proceedings of the IEEE Vol. 87, No. 8, pp. 1315-1326.

Zhou, H. H., Obuchowski, N. A., McClish, D. K., 2002: Statistical methods in diagnostic medicine, Wiley & Sons, New York.